Serial No. 10/038,984
Group No. 1635
Confirmation No. 9705

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: Not classified

A2

(11) International Publication Number: WO 90/12488

(43) International Publication Date: 1 November 1990 (01.11.90)

(21) International Application Number: PCT/US90/01717

(22) International Filing Date: 4 April 1990 (04.04.90)

(30) Priority data:
311,217 5 April 1989 (05.04.89) US
Not furnished 9 March 1990 (09.03.90) US

(71) Applicant: THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, U.S. DEPARTMENT OF COMMERCE [US/US]; 5285 Port Royal Road, Springfield, VA 22161 (US).

(72) Inventors: WICKNER, Reed, B. ; 9010 Bradgrove Drive, Bethesda, MD 20817 (US). DINMAN, Jonathan, D. ; 8720 Drexel Hill Place, Gaithersburg, MD 20879 (US). ICHO, Tateo ; 2-28-Sonechi Chofu, Tokyo 182 (JP).

(74) Agents: HOLMAN, John, Clarke et al.; Fleit, Jacobson, Cohn, Price, Holman & Stern, The Jenifer Building, 400 Seventh Street, N.W., Washington, DC 20004 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: A CLONE OF DOUBLE-STRANDED RNA VIRUS APPLIED TO ANTIBODY PRODUCTION, STUDY OF RETROVIRUS-LIKE FRAMESHIFTING AND PRODUCTION OF PROTEINS IN YEAST

(57) Abstract

A recombinant DNA molecule is described which is useful for determining the effect of an agent on translational frameshifting that is directed by a viral nucleotide sequence as a means of regulation of viral gene expression. This DNA molecule comprises an indicator gene for detection of frameshifting during translation and a DNA segment that encodes that portion of the double-stranded RNA genome of the L-A virus of yeast that directs translational frameshifting between the reading frame for the major coat protein and the reading frame for the RNA polymerase.

+ See back of page

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Monaco
AU	Australia	FI	Finland	MG	Madagascar
BB	Barbados	FR	France	ML	Mali
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GR	Greece	NL	Netherlands
BJ	Benin	HU	Hungary	NO	Norway
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo			SN	Senegal
CH	Switzerland	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
DE	Germany, Federal Republic of	LK	Sri Lanka	TG	Togo
DK	Denmark	LU	Luxembourg	US	United States of America

A CLONE OF DOUBLE-STRANDED RNA VIRUS APPLIED TO ANTIBODY PRODUCTION, STUDY OF RETROVIRUS-LIKE FRAMESHIFTING AND PRODUCTION OF PROTEINS IN YEAST

This is a continuation-in-part of the pending application Serial Number 07/311,217 filed April 5, 1989, which is a continuation-in-part of the pending application Serial Number 07/169,486 filed March 17, 1988.

BACKGROUND OF THE INVENTION

Field of the invention

The present invention relates to a DNA segment that encodes a double-stranded RNA genome of a yeast virus. More particularly, the present invention is related to a cDNA clone of the genome of the L-A virus, or a unique portion thereof, and various applications of this clone, including a method for determining the effect of an agent on translational frameshifting directed by a viral nucleotide sequence.

Background

Viruses use a wide variety of strategies to replicate and differentially express the proteins they encode. Retroviruses, double-stranded RNA (dsRNA) viruses and (+) strand RNA viruses including picornaviruses and togaviruses, have in common their use of genomic (+) strand RNA as a template for replication and as a message for protein synthesis. DNA viruses and (-) strand RNA viruses use splicing of mRNA to produce, in different amounts, proteins sharing part of their sequence (Livingston et al., 1985, Replication of Papovaviruses. In Virology, B.N. Fields, ed., New York: Raven Press, pp. 393-410; Horwitz, 1985, Adenoviruses and their replication. In Virology, B.N. Fields, ed. New York: Raven Press, pp. 433-476; Kingsbury, 1985, Orthomyxo- and

Paramyxoviruses and their replication. In Virology, B.N. Fields, ed., New York: Raven Press, pp. 1157-1178). However, splicing is unknown among (+) strand RNA viruses and dsRNA viruses, probably because the spliced RNA would be packaged and replicated leading to the accumulation of defective viruses. Retroviruses splice their genomic RNA to make the env protein, but the spliced RNA lacks the Psi sequence necessary for packaging (Watanabe et al., 1983, Mol. Cell. Biol. 3:2241-2249; Mann et al., 1983, Cell 33:153-159; Markowitz et al., 1988, J. Virol. 62:1120-1124). In addition, retroviruses use ribosomal frameshifting to make a large amount of the gag protein and a small amount of the gag-pol fusion protein (Jacks et al., 1988, Cell 55:447-458; Jacks and Varmus, 1985, Science 230:1237-1242; Kramer et al., 1986. HTLV-III gag protein is processed in yeast cells by the virus pol-protease. Science 231:1580-1584). Murine leukemia virus (Varmus, 1988, Science 240:1427-1435) and two α viruses (Strauss et al., 1983, Proc. Natl. Acad. Sci. USA 80:5271-5275) use nonsense suppression in the same way. The rates of transcription and translation of different reovirus dsRNA segments vary over a range of more than 20-fold to give overall rates of expression that vary over 400-fold (Joklik, 1981, Microbiol Rev. 45:483-501).

L-A is a dsRNA virus of the yeast Saccharomyces cerevisiae. L-A replicates by a conservative mechanism, with (+) strands made by transcription of dsRNA and (-) strands then made by copying the (+) strands to form dsRNA again

(sequential synthesis) (Newman et al., 1981, J. Virol 38:263-271; Sclafani et al., 1984, Mol. Cell. Biol. 4:1618-1626; Newman et al., 1986, The replication of dsRNA. In Extrachromosomal Genetic Elements in Lower Eukaryotes, R.B. Wickner, A. Hinebusch, A., M. Lambowitz, I.C. Gunsalus, A. Hollaender, eds., New York: Plenum Press, pp. 173-187; Fujimura et al. 1986, Proc. Natl. Acad. Sci. USA 83:4433-4437). (+) single-stranded RNA (ssRNA) is packaged in a new coat to form new viral particles (Fujimura et al., 1986, supra; Fujimura et al., 1987, Mol. Cell. Biol. 7:420-426). Both (+) and (-) strands are synthesized within viral particles, and because the size of the heads appears to be designed to contain one L-A molecule (4.6 kb), dsRNAs less than half the size of L-A replicate within the head until they fill it with multiple copies.

The 4.6 kb L-A dsRNA genome with a coding capacity of about 180 kDa encodes both the 80 kDa major coat protein (over 100 molecules per viral particle) and a 180 kDa minor viral protein (about 1 molecule per particle). The 180 kDa protein shares immunodeterminants with the major coat protein and, unlike the major coat protein, has ssRNA-binding activity which could play a role in both packaging and replication. L-A carries several genetic activities, called [HOK], [NEX], [EXL] and [B] defined by the interactions of L-A with M dsRNA (see review by Wickner, 1986, Ann. Rev. Biochem. 55:373-395; Uemura et al., 1988, Mol. Cell. Biol. 8:938-944). M_1 is a dsRNA that encodes a secreted toxin and immunity to that toxin (Bussey, 1988, Yeast 4:17-26) and uses the

same particles as L-A for replication. Although both L-A and M_1 are affected by a variety of chromosomal genes, only three of the genes required for the maintenance of M_1 (MAK3, MAK10, and PET18) are necessary for the maintenance of L-A (see review by Wickner, 1986, supra). A deletion mutant of L-A, called X dsRNA, requires most, if not all, of the MAK genes (Esteban et al., 1988, J. Virol. 62:1278-1285).

In order to determine how both the 80 kDa and 180 kDa viral proteins can be encoded by the 4.6 kb L-A molecule and to study their differential expression, structure, function, and relation to the genetic activities of L-A, it was desirable to clone and sequence the L-A genome, such a clone not previously being available.

SUMMARY OF THE INVENTION

It is, therefore, an object of the present invention to provide a substantially complete cDNA clone of the L-A viral genome of S. cerevisiae.

It is another object of the present invention to provide a yeast test system for identifying factors influencing or controlling the -1 ribosomal frameshifting event that occurs during gag-pol fusion protein synthesis in retroviruses including HIV.

It is a further object of the present invention to provide a new yeast-based vector for preparing particles-immobilized antigens for inducing an immune response in a responsive host.

It is yet another object of the present invention to provide a stable, recombinant cDNA

clone of L-A at a copy number of at least about 10,000 per cell to produce a desired protein in large amounts.

Other objects and advantages of the present invention will become evident from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects, features and many of the attendant advantages of the invention will be better understood upon a reading of the following detailed description when considered in connection with the accompanying drawings wherein:

Figure 1 shows a summary of the overall structure of the L-A genome. The two open reading frames, ORF1 and ORF2 encode the two products of the L-A genome, the major coat protein of the viral particles and the chimeric protein having a major coat protein domain and a single-stranded RNA binding-RNA polymerase domain.

Figure 2 shows the complete sequence of the L-A genome and specific features thereof, including possible frameshifting sequences near base 2000.

Figure 3 shows the similarity between the sequences of HIV-1 and RSV that are known to produce ribosomal frameshifting and sequences in L-A in the 130 base region of overlap of ORF1 and ORF2.

Figure 4 shows the homology between L-A's ORFs and the RNA-dependent RNA polymerase genes of a number of RNA viruses that allowed the identification of the 180 kDa protein as the L-A RNA polymerase.

Figure 5 shows the schematic design of a frameshift vector based on the frameshift sequences of the L-A virus.

Figure 6 shows constructs useful for producing particles carrying multiple copies of a protein antigen on its surface for inducing an immune response.

Figure 7 shows the use of a vector to support the replication and increase the copy number of the X dsRNA-based expression vector.

Figure 8 shows the structure of the L-A mRNA in the frameshift region that is involved in the "simultaneous slippage" mechanism, including the "slippery sequence", stem-loop and "pseudoknot" elements, and the use of these elements in combination with the beta-galactosidase indicator gene to provide a system for detection of frameshifting efficiency.

Figure 9 shows the sequences of three vectors for frameshift detection, pF7, pF8 and pF9. pF9 is the 0 frame vector having two stop codons before the beta-galactosidase gene in pF9 to provide a control for the level of a downstream start producing beta-galactosidase. pF8 is the -1 frame vector which detects the level of frameshifting into the -1 frame. pF7 is the +1 vector which detects shifting into the +1 frame. pT126 is a 0 frame (in-frame) vector with no stop codons which measures the amount of beta-galactosidase made from the promoter-translational start site combination of these vectors.

Figure 10 shows the minimal sequence of the vector for the assay of frameshifting in the region from the PGK1 UAS-promoter through the

artificial translation start site, the shifting sequence derived from L-A and the beta-galactosidase indicator gene. The PGK1 UAS-promoter was selected for reasons of convenience, but it will be recognized by one skilled in the art that any other yeast promoter, such as the GAL1 or PYK1 promoters, for example, can be used to provide transcription of the frameshifting region and indicator gene. The vector also includes the TRP1 gene and the 2 micron DNA replicon as well as sequences from pBR322 (not shown) which are used according to methods well known in the art for providing replication functions for the vector in bacterial and yeast host cells.

DETAILED DESCRIPTION OF THE INVENTION

The above and various other objects and advantages of the present invention are achieved by a cDNA clone of the L-A virus genome of S. cerevisiae.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. Unless mentioned otherwise, the techniques employed herein are standard methodologies well known to one of ordinary skill in the art.

The term "substantially complete" cDNA clone, as used herein, means that the clone has the complete nucleotide sequence of the L-A viral genome except for a few bases at each end which, however, are easily repaired by using oligonucleotide mutagenesis.

The term "modified" cDNA as used herein means (1) modified to include the sequence encoding the protein whose production is desired; (2) modified to place the appropriate region with this encoding sequence downstream from a suitable promoter and upstream from a suitable transcription terminator; (3) modified in a suitable region to optimize the amount of transcript production by L-A (with the protein coding sequence); (4) modified to optimize the level of translation of the mRNA produced by L-A; (5) modified to result in secretion of the synthesized protein; (6) modified to include a multiple cloning site in an appropriate region, and the like.

One skilled in the art will understand that the term "unique" with respect to an amino acid sequence means a sequence consisting of about four to six amino acids which is not found in another protein. Similarly, the term "unique" with respect to a nucleotide sequence means a sequence consisting of about 12 to 18 nucleotides which is not found in another polynucleotide.

MATERIALS AND METHODS

Materials

M-MLV reverse transcriptase, RNase H, and DNA polymerase I were obtained as a kit from

Bethesda Research Laboratories. Exonuclease III and S1 nuclease were from Promega Biotec. Modified T7 DNA polymerase (Sequence) and T4 polynucleotide kinase were from United States Biochemical. Other enzymes and substrates were obtained from Bethesda Research Laboratories, New England Biolabs, New England Nuclear, LKB-Pharmacia or Boehringer Mannheim.

Saccharomyces cerevisiae strain RE59 (a arg1-1 ski2-2 L-A-HNB [D] L-BC W) was used as a source of VLPs. This strain contains a variant of L-A, L-A-HNB, which results in high levels of L-A viral particles (Uemura et al., 1988, Mol. Cell. Biol. 8:938-944). This strain has no M_1 killer dsRNA, which, if present, would strongly repress L-A replication, and the chromosomal ski2-2 mutation also contributes to the higher yield of VLPs. Escherichia coli strains, DH5α(F', endA1, hsdR17(r^-,m^+).supE44, thi-1, λ-, recA1, gyrA96, relA1, Δ(argF-lacZYA)U169, ϕ80lacZΔM15) and DH5αF', were from Bethesda Research Laboratories. Plasmids SK^+, SK^-, and M13 helper phage, R408, were from Stratagene.

Construction of cDNA Clone

L-A cDNA was synthesized using (+) strand message of L-A, synthesized *in vitro*, as a template. L-A VLPs were prepared from strain RE59 as described by Fujimura et al., 1986, Proc. Natl. Acad. Sci. USA 83:4433-4437, except that the concentration of NaCl in the extraction buffer (buffer A) was raised to 500 mM to stabilize particles, and CsCl density gradient centrifugation was performed twice. L-A (+)

strand RNA was synthesized from the purified VLPs as described by Welsh et al., 1980, Nucleic Acids Res. 8:2349-2363. To obtain the highest yield of the product per template in a 200 µl reaction volume, each substrate (NTP) concentration was raised to 1.5 mM and the $MgCl_2$ concentration was adjusted to 10 mM. After 90 minutes incubation at 30°C, the reaction was stopped by phenol, and then standard phenol-chloroform extraction followed. After ethanol precipitation, the mixture of L-A ssRNA and dsRNA was used directly as a template for the cDNA synthesis without further manipulation. The presence of dsRNA interfered neither with cDNA synthesis nor with the ligation reaction.

Two synthetic primers, GCATATGGGTAATTCCCATTATCTTTTTGGC (PRIMER I) and GAAAAATTTTTAAATTCATATAACTCCCC (PRIMER II), were used to synthesize L-A cDNA. These primer sequences were based on the 3' and 5' terminal sequences of L-A dsRNA (Thiele et al., 1984, Mol. Cell. Biol. 4:92-100). The 5'-termini of these primers were phosphorylated by T4 polynucleotide kinase. The first strand was synthesized with M-MLV reverse transcriptase following the manufacturer's protocol, except that oligo dT was replaced by the primer I. After the first strand synthesis, the primer II was included in the reaction mixture, annealed, and the second strand synthesis was performed using DNA polymerase I with RNaseH. This use of the second primer allowed us to generate cDNA having an intact 5'-end. No labeled nucleotide was used at any point in the reaction process. The reaction was

terminated by a phenol extraction and the full size cDNA which runs at the same position as dsRNA was purified from a preparative agarose gel using DEAE-cellulose paper. The size fractionated cDNA was ligated into the *Sma*I site of the multiple cloning site of a Bluescript vector, SK⁺. The ligated DNA mixture was transformed into frozen competent cells of *E. coli* strain DH5α prepared by the method of Hanahan, 1983, J. Mol. Biol. 166:557-50. Transformants were selected on X-gal ampicillin plates and white and pale blue colonies were screened for their plasmid DNA inserts.

Initially 108 transformants were screened for their insert DNA size, by digesting mini-prep DNA with *Hind*III and *Bam*HI. Among them, 39 clones were close to the full size. These clones were further tested by digesting the plasmids by *Eco*RI, *Eco*RV, and *Bam*HI, and comparing the size of the fragments derived from the 5′ and 3′ ends of the L-A cDNA. The digested fragments were transferred from an agarose gel to two pieces of nitrocellulose paper and hybridized with ^{32}P end-labeled primers I and II, respectively. Among those which showed strong hybridization to both primers were L03, L05, and L28.

DNA Sequencing of L-A cDNA Clone

The plasmids, L03 and L05, have L-A cDNA inserts whose orientation is such that the 5′ end of the L-A (+) strand sequence is next to the universal primer site in the SK⁺ vector, whereas the insert in L28 has the opposite orientation. These plasmids were digested at the *Hind*III and ApaI sites in the multiple cloning site of the SK⁺

vector, and a series of deletions were generated using exoIII and S1 nuclease (Henikoff, 1987, Methods Enzymol. 155:156-165). ssDNAs were prepared following the protocol supplied by Stratagene with the following modifications. Fresh colonies of strain DH5αF' carrying a plasmid were inoculated into 2 ml of 2xYT medium in a 50 ml disposable plastic tube, cultured at 37°C for 2 hours, and M13 helper phage, R408, was infected at an moi of about 10. After vigorous shaking for two hours at 37°C, the temperature was dropped to 32°C, and the shaking was continued overnight. Although the reduction of the temperature is not essential, this method gave a consistent yield of rescued ssDNAs in a variety of deletion mutants constructed. Since F-pili are dissociated at the reduced temperature (Messing, 1983, Methods Enzymol. 101:20-78), this modification prevented reinfection by the helper phages which could eventually lyse the cells, especially those cells which have lost the plasmid during the overnight cultivation.

For each plasmid, about 50 deletions were sequenced using Sequenase following the manufacturer's protocol. The average reading was 400 bp. For clones in the SK' vector, the M13 -20 17 mer was used as a primer. Also, *BamHI/HindIII* fragments from 6 cDNA clones were recloned into the SK' vector, and the opposite ends were sequenced using a reverse sequencing primer. The DNA sequences were assembled and analyzed using programs of UWGCG (Devereux et al., 1984, Nucleic Acids Res. 12:387-395), IDEAS, STADEN, and PIR

included in the "Analysis" library at the National Cancer Institute Computer Center at Frederick, Maryland. Homology was also examined using the FASTA (Pearson et al., 1988, Proc. Natl. Acad. Sci USA 85:2444-2448) and FPAT programs provided by David Lipman, Chuck Buckler and William Pearson.

A deposit of the cDNA clone of the L-A dsRNA has been made at the American Type Culture Collection, Rockville, Maryland, on February 10, 1989 under accession number 67,888. The deposit shall be viably maintained, replacing if it becomes non-viable, for the life of the patent or a period of 30 years from the date of the deposit, or for 5 years from the last date of request for a sample of the deposit, whichever is longer, and made available to the public without restriction in accordance with the provisions of the law. The Commissioner of Patents and Trademarks, upon request, shall have access to the deposit.

Among various utilities of the clone of the present invention, at least the following shall be noted.

1. To identify factors influencing the ribosomal frameshifting that retroviruses, including HIV, use to make the gag-pol fusion protein which is vital to viral replication.

2. To prepare particles that have a particular antigen protruding from multiple sites on the particle surface for use in immunization or in biological and immunological tests.

3. To provide an RNA expression vector and a helper for such an RNA expression vector. These utilities are now described.

Use of cDNA clones of L-A in examining factors affecting ribosomal frameshifting.

A study of the DNA sequence of the cDNA clone of the present invention reveals that the 4579 bp L-A has two open reading frames (ORFs). ORF1 of 680 amino acids extends from base 30 to base 2072 of the L-A(+) strand and encodes the major coat protein. A second ORF of 868 amino acids (ORF2) extends from base 1940 to 4546 and encodes part of the ssRNA-binding protein. The other part of this 180 kDa ssRNA-binding protein is encoded by ORF1. Fusion of ORF1 and ORF2, apparently by a -1 translational frameshift, produces the complete 180 kDa protein. A site similar to the site of frameshifting in HIV, RSV and other retroviruses is found within the 130 bp overlap of ORF1 and ORF2 (Fig. 3). ORF2 also contains a sequence characteristic of the RNA-dependent RNA polymerases of several picorna- and togaviruses (Fig. 4).

Retroviruses, including HIV, use ribosomal frameshifting to make a large amount of the gag protein (the major coat protein of the viral cores) and a small amount of the gag-pol fusion protein. The gag-pol fusion protein contains a number of domains, such as a domain that is identical to the gag protein itself, a domain that has the reverse transcriptase activity, and other protease, integrase and RNase H domains. This fusion protein is essential for the retroviral replication cycle and, therefore, interference with the process of ribosomal frameshifting, either by decreasing its efficiency or increasing its efficiency, would be expected to

block the retroviral replication cycle. At present, -1 ribosomal frameshifting is not known to be used by eukaryotic cells for their own purposes. Most known cases are in retroviruses, while one is in a coronavirus, one in the L-A double-stranded RNA virus of yeast (described herein) and several suspected cases in (+) strand RNA viruses. This means that, based on evidence to date, a drug specifically affecting ribosomal frameshifting would be effective against various viral infections, mostly retroviruses, but would not adversely affect the host. In order to detect and develop antiviral agents which act by influencing frameshifting, it is necessary to have a simple system in which the effect of various drugs, host factors or other perturbations on the efficiency of ribosomal frameshifting can be easily tested and determined.

The retroviruses must make the host ribosomes carry out frameshifting from the upstream gag reading frame to the downstream pol reading frame. This must be carried out while the ribosomes are in the rather small region (201 bp in the case of HIV) in which the two ORFs overlap. In all known cases, mammalian retroviruses carry out a shift of -1 base in the overlap region.

As mentioned above, ORF1 of L-A encodes the major coat protein of the viral particles in which L-A is found in the cell (like gag of retroviruses). ORF2 encodes a part of the 180 kDa protein with homology to RNA-dependent RNA polymerases that has single-stranded RNA binding activity. Like the gag protein of retroviruses, the L-A major coat protein encoded by ORF1 is made

in large amounts. Like the gag-pol fusion protein of retroviruses, the ORF1-ORF2 fusion protein (the 180 kDa protein) is made in small amounts. The mechanism appears to be same. The shift from ORF1 to ORF2 must occur in the 130 bp region of overlap between ORF1 and ORF2. The exact site of the shift in HIV, RSV and in several other retroviruses has been precisely determined and has been shown to occur at the sites shown in Fig. 3 for HIV and for RSV. The essential elements of this site are the sequences, 5′ UUUUUUA 3′ for HIV or 5′ AAAUUUA 3′ for RSV, followed closely by an inverted repeat sequence that most likely functions to slow down the ribosomes. As shown in Fig. 3, a similar sequence was found in L-A in the 130 bp overlap region. The reading frame in this homologous region is the same as that in HIV where it frameshifts. The ORF1 frame in this GGGUUUA sequence is such that a -1 basé "simultaneous slippage" (Jacks et al., 1988, Nature 331:280-283; Jacks et al., 1988, Cell, 55:447-458) of peptidyl $tRNA^{Gly}$ bound to the GGU and leucyl $tRNA^{Leu}$ bound to the UUA leaves the non-wobble bases of each tRNA anticodon still paired to the mRNA. This mechanism has been demonstrated by Jacks et al., supra., for RSV, and is strongly suggested for other retroviruses which frameshift. The L-A double-stranded RNA virus makes its gag-pol fusion protein by the same mechanism (see Fig. 8) as that used by retroviruses, namely, a -1 ribosomal frameshifting event by the "simultaneous slippage" mechanism described for Rous Sarcoma Virus (Jacks and Varmus, 1985, Science 230:1237-1242; Jacks and Varmus, 1988, Nature 331:280-283). The general

structure and sequence pattern at the site of the frameshift is the same as that used by retroviruses, namely a short "slippery sequence" (GGGUUUAGC for L-A) upstream from a stem-loop structure which has sequences complementary to a downstream region such that a secondary structure known as a "pseudoknot" can be formed in the mRNA (Fig. 8).

Since ribosomal frameshifting is an interaction of the ribosome and the mRNA, it is reasonable to use any message construct to examine its mechanism whether or not that construct comes from a retrovirus or not. At the time the ribosomes are frameshifting on HIV message, they have no way of knowing whether the RNA came from a retrovirus or not. The yeast system of the present invention provides easy genetic manipulability for host genes, viral components and clones.

In order to make a construct, the region at which the frameshifting occurs is placed upstream of an indicator gene for detection of frameshifting, for example, the beta-galactosidase (lacZ) gene of *E. coli*, which can be detected by an indicator dye, or a gene such as LEU2 or URA3 or LYS2 or CAN1 for which cells having increased or decreased expression can be selected (Fig. 5). For example, cells expressing the products of the LEU2, URA3, or LYS2 genes, which are involved in intermediary metabolism, can be selected by omission of a required nutrient from the culture medium. In addition, under other conditions, cells which are not expressing the same URA3, LYS2 or CAN1 genes can be selected by addition to the

medium of a metabolic inhibitor for which toxicity is mediated by the URA3 or LYS2 gene product.

In construction of the vectors for frameshift detection, the reading frame is adjusted so that only if frameshifting occurs will the downstream signal gene be expressed. One or more of these constructs (Fig. 5) are introduced into yeast strains and mutants are isolated in genes that affect frameshifting efficiency by either increasing or decreasing it. Drugs are screened for their effects on frameshifting using cells containing these "frameshifting vectors". A decrease in frameshifting would interfere with retroviral replication by interfering with the supply of reverse transcriptase. Monitoring the effects on more than one construct makes the screening of effects peculiar to one signal gene a simple matter. Also, other control constructs are prepared in which the signal gene is in frame with the upstream region so that no frameshift is needed for expression. Yet other control constructs have the signal gene in the +1 frame.

A set of vectors has been constructed with the frameshift site of the L-A virus downstream of a yeast constitutive promotor and upstream of the lacZ gene of *E. coli* in each of the three possible reading frames (Figures 9 and 10). Stop codons are placed such that only a frameshift in the region of the overlap of the open reading frames can generate active beta-galactosidase. If 100% is the level of beta-

galactosidase produced from the vector with the lacZ gene in frame with the promotor, then the -1 frame vector produces 2% beta-galactosidase activity and the +1 frame vector produces only about 0.05% beta-galactosidase. Thus, this structure produces specifically -1 ribosomal frameshifting. It has been confirmed by subcloning experiments that this specific region (L-A bases 1950 to 2028 from the 5' end of the (+) strand) is actually responsible for the frameshifting. Based on the current understanding of the mechanism of frameshifting in the L-A virus and retroviruses, the minimal sequence required to direct translational frameshifting in this region begins at about the start of the slippery site (i.e., at the position numbered 1958) and ends with the last base of the 3' pseudoknot region (at position 2024), inclusive, from the 5' end of the (+) RNA strand, as shown in Figure 10.

This requirement for the secondary structure downstream from the L-A slippery site stands in contrast to a report about the sequence requirements for frameshifting in HIV in which Wilson et al., (1988. HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell 55:1159-1169) teach that frameshifting in HIV is directed by the slippery site alone and is not dependent on downstream stem-loop structures. In the present system, it was found that deletion near the frameshifting region on the 5' side of the slippery site, up to position 1956, has no effect on frameshifting. Substitution of another base for one or two of the first three G's in the

slippery site (GGGUUUAGG) interferes with frameshifting, but frameshifting is restored if all three of these G's are replaced with the same base. On the 3' side, deletions up to the 3' end of the 3' pseudoknot sequence have no effect, whereas base substitutions in the 3' pseudoknot interfere with frameshifting. Furthermore, base changes in the 5' pseudoknot sequence that render this sequence noncomplementary to the 3' pseudoknot sequence also interfere with frameshifting. Accordingly, it will be recognized by one of ordinary skill in the art of genetic engineering that certain substitutions of base sequences can be made in this L-A virus frameshifting region without affecting the frameshifting function and, therefore, without departing from the spirit and purview of this application and scope of the appended claims. Specifically, in addition to the changes cited above, changes in the stem-loop structure which preserve its inherent complementarity, its approximate size, and the relative positions of the pseudoknot elements can be made without materially affecting the frameshifting function.

The vectors described above can now be used for the rapid screening of drugs potentially affecting the efficiency of -1 ribosomal frameshifting using an embodiment of the invention comprising a microbial screening assay on semisolid culture medium. Beta-galactosidase produced *in vivo* can be assayed on petri plates by incorporating X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) in the solid medium on which the cells carrying the vector are grown by

methods well known in the art (T.J. Silhavy, Berman ML, and Enquist LM: Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring, Harbor, NY, 1984). The beta-galactosidase produced releases a blue dye from the X-gal, and the intensity of the color is an extremely sensitive indication of the efficiency of -1 ribosomal frameshifting. This color intensity is capable of reproducibly indicating differences in beta-galactosidase production as low as 2 fold, but it is also capable of indicating large difference as well. The low (2%) level of enzyme produced by the -1 ribosomal frameshifting vector results in a light blue color while the 100% vector (0 shift) gives very dark blue and the 0.05% level of the +1 shift vector gives white colonies.

These vectors have been used to isolate mutants in chromosomal genes affecting the efficiency of frameshifting. Starting with the light blue wild-type host carrying the -1 vector, white and dark blue mutants were picked. The mutants were assayed for beta-galactosidase directly and the white mutants always had reduced amounts while the dark blue mutants always had increased amounts of enzyme activity, confirming the accuracy of the plate assay.

The identical vector in the same strain, or a strain with increased permeability to drugs, is used to screen drugs on plates. Drug solutions are applied directly to a newly seeded lawn of the vector-containing strain on an X-gal plate. The effect of the drug on the blue color development indicates the effect of the drug on -1 ribosomal

frameshifting. Gradient plates with an under layer of agar containing the drug could also be used. Potentially active drugs could also be tested for effects on cells carrying the 0 frame or +1 frame vectors or no vector, as controls for effects on other steps in the degradation of X-gal. Determination of effects on growth of the cells provides a simple first screen for drug toxicity.

Using such a test, thousands of drugs can be screened quickly and cheaply since solid medium for yeast growth is very inexpensive and the manipulations are simple. It is expected that an agent which showed either a substantial increase or decrease in ribosomal frameshifting efficiency in the present system would adversely affect retroviral reproduction and thus be useful as an antiviral agent.

Use of L-A clones to make antigen-carrying particles for Immunization

Particle-bound antigens have long been known to be more antigenic than soluble antigens (Adams et al., 1987, Nature 329:68-70). Using the process of Adams et al., gene fusions or ORF1 encoding the major coat protein of L-A are made with the gene of the protein antigen against which immunization is desired. The fusion is introduced into yeast and expressed. The particles formed by the major coat protein then have the protein antigen attached at multiple locations on each of the particles. Expression at various levels in the presence or absence of endogenous L-A virus can be used to produce particles having various

numbers of antigen side chains per particle. The DNA sequence for the protein antigen is introduced at various sites within the major coat protein gene to determine which sites give the most immunogenic particles. The particles can be purified easily using their physical properties. The particles can then be inoculated to induce the desired immune response.

Use of L-A cDNA clones to express and produce a desired protein and as a helper for an X cDNA-based RNA vector.

In the replication cycle of the L-A virus, the viral (+) strands are an intermediate. Viral (+) strands are synthesized by viral particles that contain the viral double-stranded RNA genome and extruded from the viral particles. They then serve a dual role;

(I) These (+) strands are the mRNA that is used by the ribosomes to make the two viral proteins needed for viral packaging and replication, namely, the major coat protein (80 kDa) and the major coat protein-RNA polymerase fusion protein (180kDa).

(II) The viral plus strands are encapsidated by the newly synthesized coat proteins to form new virus particles.

This natural viral replication cycle is exploited to synthesize abundant amounts of a desired protein. The L-A cDNA clone can also be modified to encode the desired protein instead of, or in addition to, the normally encoded proteins. The viral regions and signals necessary for the replication, transcription and packaging of the L-

A RNA molecule have been defined and delimited. These signals are completely included in the 5' most 25 bases of the (+) strand and the 3' most 500 bases of the (+) strand (Esteban et al., 1988, Proc. Natl. Acad. Sci. USA, 85:4411-4415). Replication requires the 3' terminal 30 bases and a region about 400 bases from the 3' end (the Viral Binding Sequence or the Internal Replication Enhancer). In order to produce the proteins, the region coding for the viral proteins is replaced by the gene for the desired protein. The signals for transcription and packaging and replication are left intact and the desired protein is allowed to be expressed.

Since strains having up to 20% of their total protein as major coat protein can be constructed, the transcription and translation signals normally used by the virus are quite efficient, although the method allows one to modify and improve either.

The L-A cDNA clone is altered as above to contain the desired gene and upstream of the modified L-A sequence is placed either a yeast promoter or a T7 RNA polymerase promoter. This vector is then introduced into a yeast strain that carries a normal L-A virus to supply the replication and packaging proteins. Then, the transcription of the modified L-A is turned on. The (+) strands thus produced (because they have the packaging sequences) are packaged in virus particles replicated and transcribed. The (+) strand transcripts are translated to produce the desired product.

Another use for the L-A cDNA clone in producing the desired protein is as a source of large amounts of replication proteins. The normal L-A produces the replicase (RNA-dependent RNA polymerase) as a fusion protein with the major coat protein at about 1% of the amount of coat protein alone. While this ratio may be optimal for maintenance and replication of L-A in a normal host strain, it is probably sub-optimal for maximal copy number and expression in the ski- mutant hosts that give up to 10-20% of host proteins as L-A major coat protein. The mechanism of formation of the fusion protein is ribosomal frameshifting. This is an inefficient process and can be simply bypassed by inserting one nucleotide in the region of overlap between the two open reading frames of the L-A cDNA clone. If this clone is modified so as not to have the packaging and replication signals, the transcripts will only serve as mRNA. Thus, a large amount of RNA-dependent RNA polymerase can be supplied to the replicating virus to increase its copy number and rate of transcription.

The invention is now illustrated by the following example:

Example 1:

This example shows the construction of a 'frameshift vector' using the frameshift region of the L-A viral sequence. The structure is schematically illustrated in Fig. 5. The yeast PYK 1 promoter is linked to an AUG start codon followed by the frameshift sequence from the L-A virus. The region of L-A analogous to HIV or RSV, shown in Fig. 5, is the exact region of L-A

responsible for frame shifting and was inserted during the construction. At the end of the L-A frameshift sequence is a termination codon, TAA, in the unshifted frame, so that ribosomes that have not shifted will cease translation and not have the opportunity to shift further downstream. The *E. coli* lacZ gene encoding beta-galactosidase, is placed next and is in the -1 frame relative to the AUG. This means that the ribosomes must shift back one base on the mRNA in order to properly translate this region. Control vectors with beta-galactosidase in the original "0" frame or in the "+1" frame are also made, but are not shown in Fig. 5. The vector also carried replication origins and selectable markers for yeast and *E. coli*. The vectors are transformed into a wild-type yeast strain and the levels of beta-galactosidase in the "0" frame with the "-1" frame and the "+1" frame. The amount of beta-galactosidase in a colony can also be estimated using plates containing X-gal and the blue color that results from its breakdown by the enzyme. Cells carrying the "-1" vector (the construct shown in Fig. 5) are grown on plates containing a drug to be tested for its effects on frameshifting. A change in the shade of blue can be used as a simple qualitative screen for an increase or decrease of the degree of frameshifting. Thus, hundreds of drugs can be quickly screened for such an effect. More careful quantitative assays can then be carried out on promising drugs using the simple beta-galactosidase assay of permeabilized whole cells. Since the yeast translation apparatus closely

resembles that of animal cells, this method can be used to screen and determine the direct effects of drugs on the replication of retroviruses in tissue culture or *in vivo* tests.

The same strain carrying the frameshift vector can be mutagenized and plated for single colonies. Those colonies that are more or less blue than the parent strain can then be analyzed to detect single-gene host cell mutations causing increased or decreased ribosomal frameshifting. The host genes involved can be cloned and characterized and their mammalian counterparts examined for similar effects on translation in mammalian cells.

Example 2:

To produce particles with regions of the HIV-1 p^{24} protein exposed on the outer surface, constructs such as are shown ih Fig. 6 are prepared. The L-A ORF1 encoding the major coat protein is fused in frame to the p^{24} gene. [In other constructs not shown the p^{24} sequence is inserted within the L-A ORF1 or at the N-terminus.] The expression vector carrying this construct is introduced into a yeast host. The major coat protein with the p^{24} sequences attached forms viral particles which are then purified in large amounts by conventional methods. The L-A ORF1-ORF2 fusion protein is expressed separately from a separate vector to determine whether this fusion protein is, as has been hypothesized, necessary for particle formation (Fujimura & Wickner, 1988, Cell, 55:663-671).

The purified particles are used to induce an immune response or to generate antibody reagents. The particles are also used as the antigen in tests for determining the presence of antibody in serum of patients or animals.

Example 3:

The L-A ORF1 and ORF2 are fused in an expression vector by simply inserting one base in the region of overlap between ORF1 and ORF2. This vector then expresses the viral RNA polymerase. Along with another expression vector expressing only the major coat protein, ORFs are used to supply needed products to the X dsRNA-based expression vector (described in U.S. patent application Serial Number 07/169,486, filed March 17, 1988) which has all cis sites necessary for replication, transcription and packaging. Using the cDNA clone of X, the gene for a protein whose production is desired can be inserted, introduced into yeast, regenerated as an RNA virus and expressed by the L-A encoded transcription apparatus. The cloning of the L-A genome makes possible supplying or modifying the protein components of the L-A packaging and transcription apparatus from these clones in such a way as to optimize expression from the modified X virus-vector. This is schematically shown in Fig. 7.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included

within the spirit and purview of this application and scope of the appended claims.

* * *

For purposes of completing the background description and present disclosure, each of the published articles, patents and patent applications heretofore identified in this specification are hereby incorporated by reference into the specification.

WHAT IS CLAIMED IS:

1. A recombinant DNA molecule comprising an indicator gene for detection of frameshifting during translation and a DNA segment,

wherein said DNA segment encodes that portion of the double-stranded RNA genome of the L-A virus of yeast consisting essentially of a viral nucleotide sequence that directs translational frameshifting between the reading frame for the major coat protein and the reading frame for the RNA polymerase, and

wherein said DNA segment is located upstream from said indicator gene so that said indicator gene is translated in the correct reading frame only when frameshifting occurs in the messenger RNA sequence transcribed from said DNA segment.

2. The recombinant DNA molecule according to claim 1 wherein said viral nucleotide sequence includes bases numbered 1958 to 2024, inclusive, shown in Figure 10.

3. The recombinant DNA molecule according to claim 1 wherein said indicator gene is selected from the group consisting of the beta-galactosidase (lacZ) gene of *Escherichia coli*, and the following genes of *Saccharomyces cerevisiae*: LEU2; URA3; LYS2 and CAN1.

4. The recombinant DNA molecule of claim 3, wherein said DNA molecule includes the entire nucleotide sequence shown in Figure 10.

5. A method for determining the effect of an agent on translational frameshifting, said frameshifting being directed by a viral nucleotide sequence, comprising:

producing the translation product of said indicator gene of said recombinant DNA of claim 1 in the absence of said agent;

producing the translation product of said indicator gene of said recombinant DNA of claim 1 in the presence of said agent; and

comparing the amount of said translation product produced in the presence of said agent with said amount produced in the absence of said agent, wherein a difference in said amount in the presence of said agent is indicative of the effect of said agent on said frameshifting.

6. The method of claim 5 wherein said recombinant DNA molecule is contained within a host cell.

7. The method of claim 6 wherein said host cell is a yeast cell.

8. The method of claim 7 wherein said translation product is beta-galactosidase.

9. The method of claim 8 wherein said host cell is cultured on a semi-solid medium and said amount of said beta-galactosidase produced by said host cell is determined by the intensity of color produced by a dye that is released by hydrolysis of a substrate for said beta-galactosidase.

10. The method of claim 7 wherein said translation product produced by the recombinant DNA molecule of claim 1 as a result of said frameshifting is a product for which cells having increased or decreased expression can be selected by varying culture conditions.

11. The method of claim 10 wherein said translation product is selected from the group consisting of the translation products of the following *Saccharomyces cerevisiae* genes: LEU2; URA3; LYS2 and CAN1.

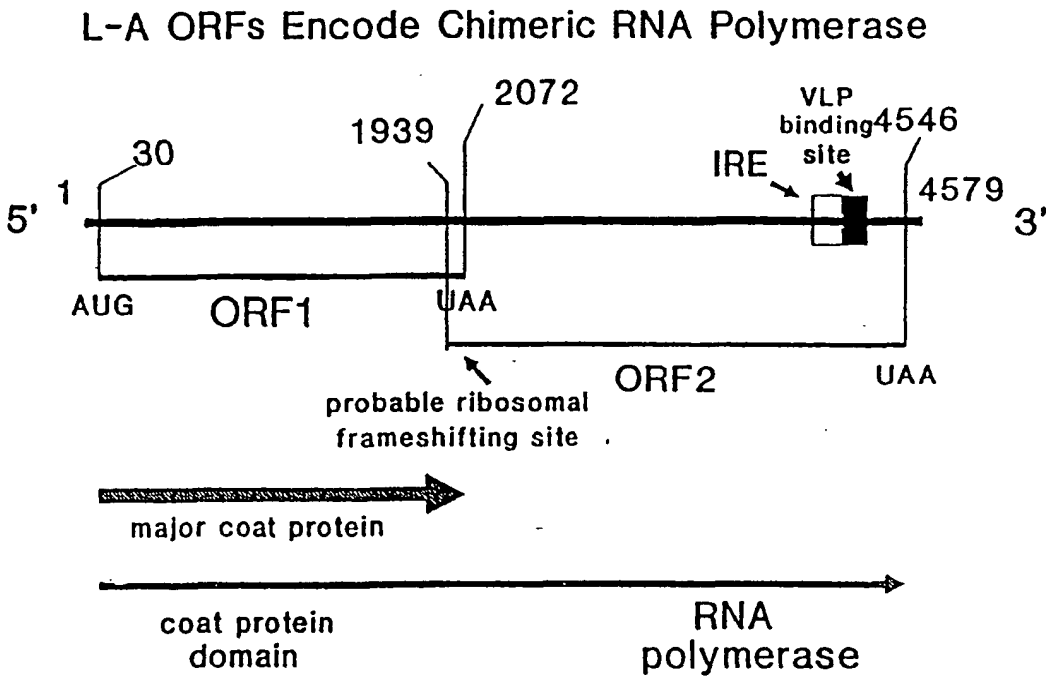

L-A ORFs Encode Chimeric RNA Polymerase
2072
1939
VLP
binding
site
4546
30
IRE
1
4579
5'
3'
AUG
ORF1
UAA
ORF2
UAA
probable ribosomal
frameshifting site
major coat protein
coat protein
domain
RNA
polymerase

Figure 2.

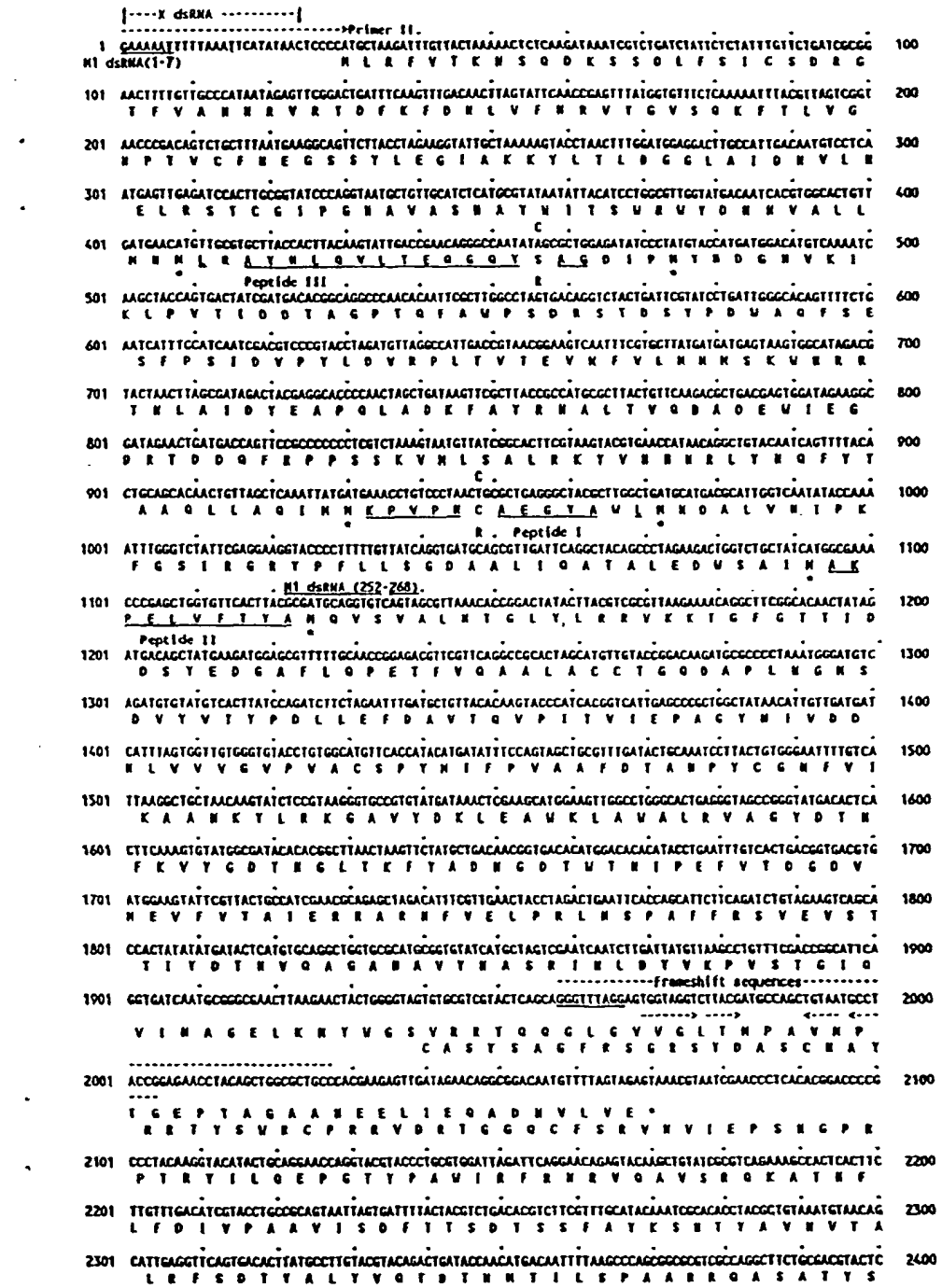

X dsRNA
Primer II.
M1 dsRNA(1-7)
Peptide III
Peptide I
M1 dsRNA (252-268).
Peptide II
frameshift sequences

Figure 2. (continued)

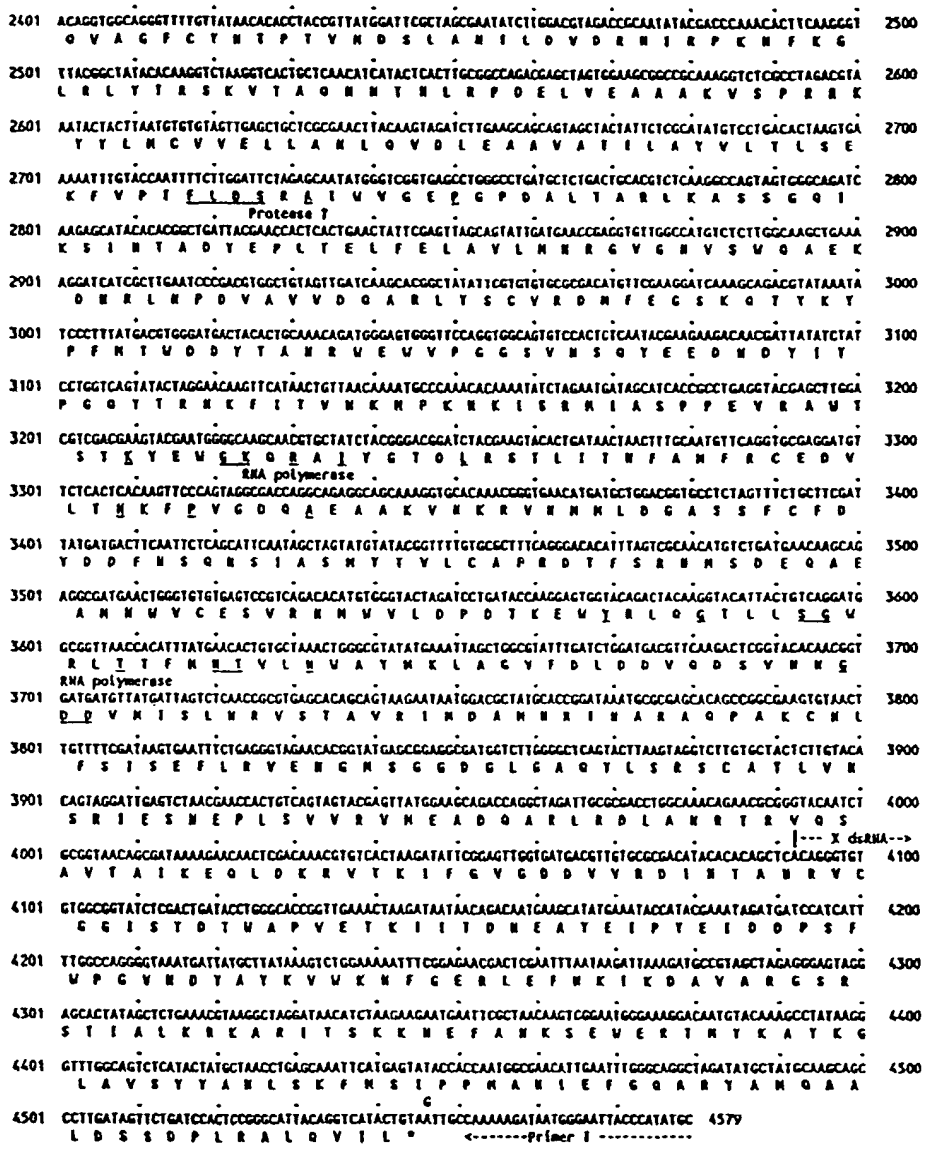
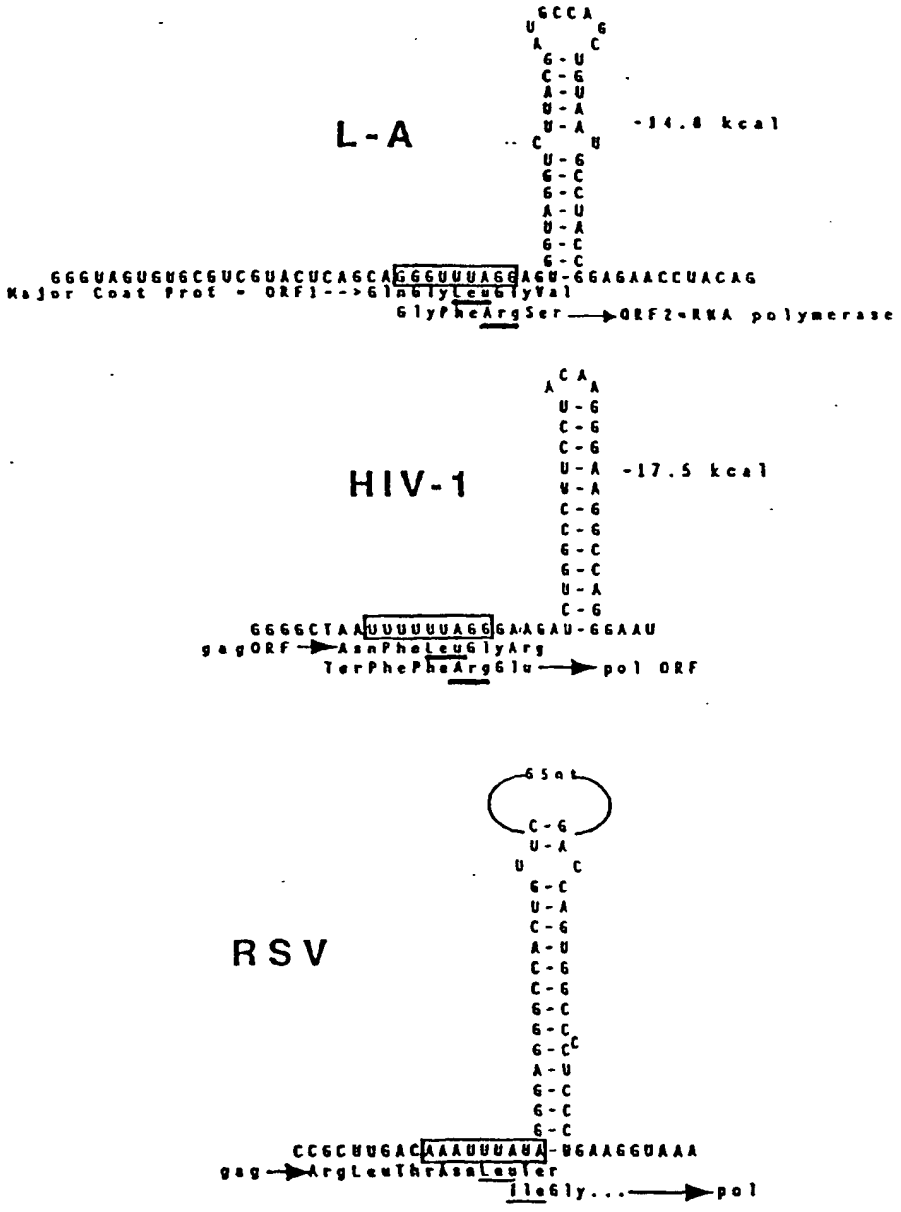


Figure 3.


L-A
-14.8 kcal
GGGUAGUGUGCGUCGUACUCAGCAGGGUUUAGGAGU-GGAGAACCUACAG
Major Coat Prot - ORF1-->GlnGlyLeuGlyVal
GlyPheArgSer
ORF2-RNA polymerase
HIV-1
-17.5 kcal
GGGGCTAAUUUUUUAGGGAAGAU-GGAAU
gagORF
AsnPheLeuGlyArg
TerPhePheArgGlu
pol ORF
RSV
CCGCUUGACAAAUUUAUA-UGAAGGUAAA
gag
ArgLeuThrAsnLeuLeu
IleGly...
pol

FMD 2024
EMC 1987
Rhino14 1876
Hepat A 1034
Polio 1903
CoxsacB3 157
L-A ORF2 413
Region I
Carnat 524
TobEch 2580
Mdlbrg 804
WNileF 3120
YellowF 3100
Rhino14 2006
Hepat A 1191
Polio 2025
Coxsac B 300
L-A ORF2 541
Region II
Region III

Figure 5.

Frameshift Vector Based on the L-A Virus

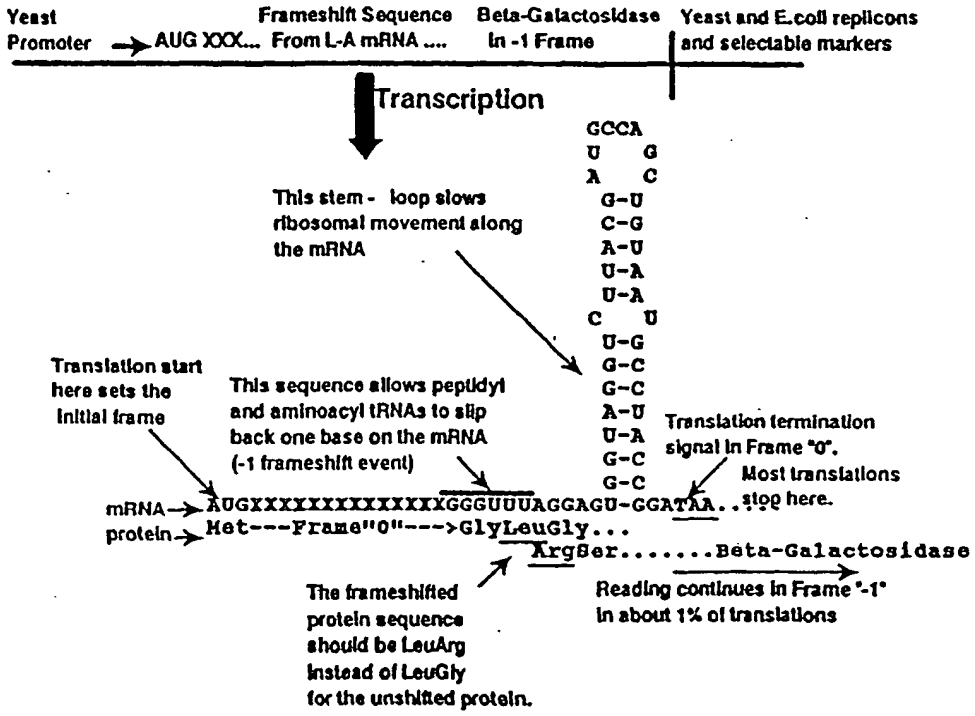


Figure 6.

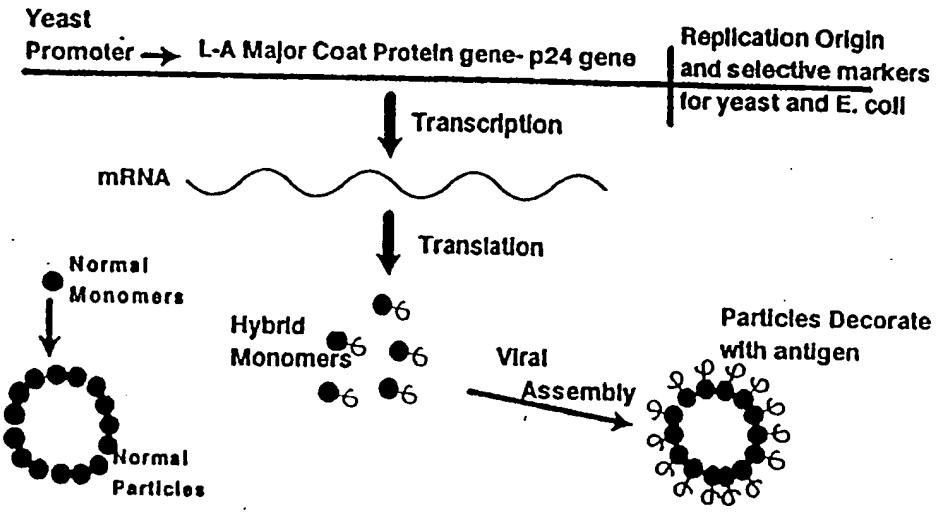

Yeast
Promoter
L-A Major Coat Protein gene- p24 gene
Replication Origin
and selective markers
for yeast and E. coli
Transcription
mRNA
Translation
Normal
Monomers
Hybrid
Monomers
Viral
Assembly
Particles Decorate
with antigen
Normal
Particles

Figure 7.

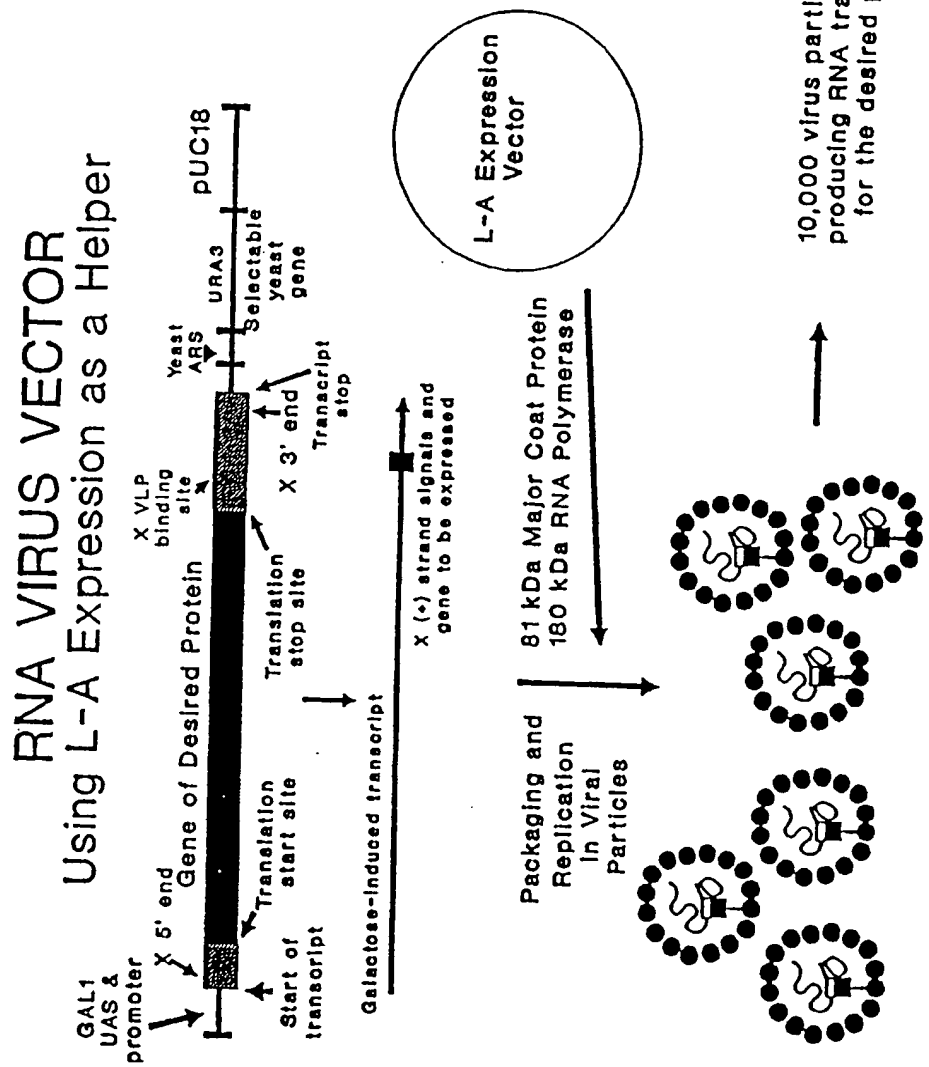
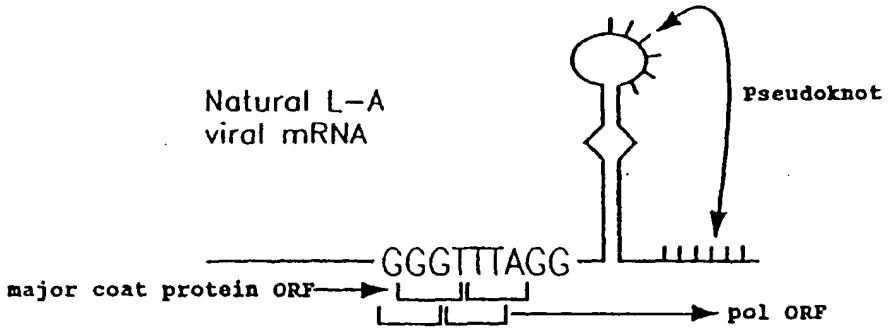
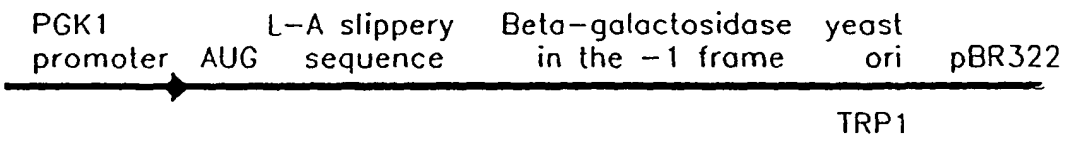

RNA VIRUS VECTOR
Using L-A Expression as a Helper
GAL1 UAS & promoter
X 5' end
Gene of Desired Protein
Translation start site
Start of transcript
X VLP binding site
Translation stop site
X 3' end
Transcript stop
Yeast ARS
URA3
Selectable yeast gene
pUC18
Galactose-induced transcript
X (+) strand signals and gene to be expressed
L-A Expression Vector
81 kDa Major Coat Protein
180 kDa RNA Polymerase
Packaging and Replication In Viral Particles
10,000 virus particles/cell producing RNA transcripts for the desired protein

Figure 8.



Frameshift Screening Vector



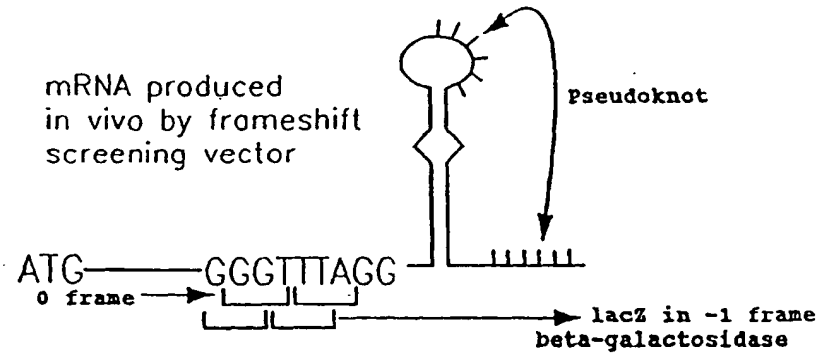
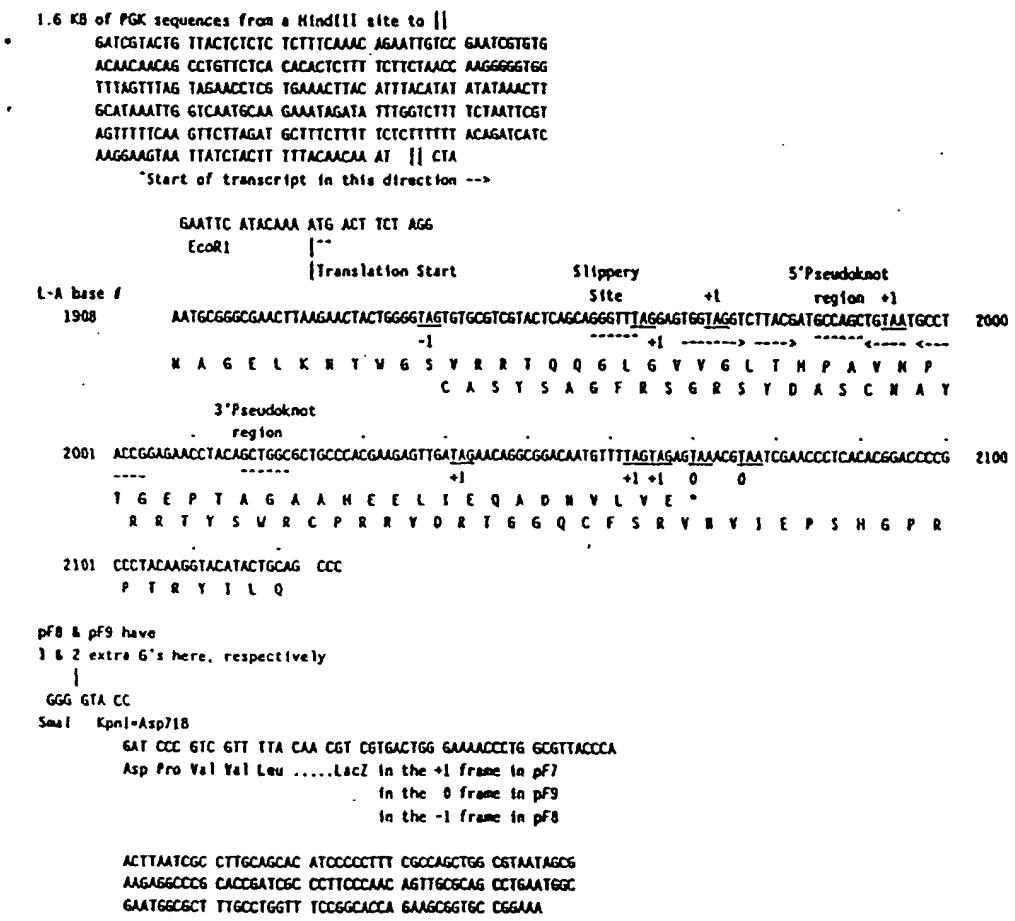


Figure 9.


1.6 KB of PGK sequences from a HindIII site to ||
GATCGTACTG TTACTCTCTC TCTTTCAAAC AGAATTGTCC GAATCGTGTG
ACAACAACAG CCTGTTCTCA CACACTCTTT TCTTCTAACC AAGGGGGTGG
TTTAGTTTAG TAGAACCTCG TGAAACTTAC ATTTACATAT ATATAAACTT
GCATAAATTG GTCAATGCAA GAAATAGATA TTTGGTCTTT TCTAATTCGT
AGTTTTTCAA GTTCTTAGAT GCTTTCTTTT TCTCTTTTTT ACAGATCATC
AAGGAAGTAA TTATCTACTT TTTACAACAA AT || CTA
*Start of transcript in this direction -->
GAATTC ATACAAA ATG ACT TCT AGG
EcoR1
Translation Start
Slippery Site
5'Pseudoknot region
+1
1908
AATGCGGGCGAACTTAAGAACTACTGGGGTAGTGTGCGTCGTACTCAGCAGGGTTTAGGAGTGGTAGGTCTTACGATGCCAGCTGTAATGCCT
2000
-1
+1
M A G E L K N Y W G S V R R T Q Q G L G V V G L T M P A V N P
C A S Y S A G F R S G R S Y D A S C N A Y
3'Pseudoknot region
2001
ACCGGAGAACCTACAGCTGGCGCTGCCCACGAAGAGTTGATAGAACAGGCGGACAATGTTTTAGTAGAGTAAACGTAATCGAACCCTCACACGGACCCCG
2100
+1
+1 +1 0 0
T G E P T A G A A H E E L I E Q A D N V L V E *
R R T Y S W R C P R R Y D R T G G Q C F S R V N V I E P S H G P R
2101
CCCTACAAGGTACATACTGCAG CCC
P T R Y I L Q
pF8 & pF9 have
1 & 2 extra G's here, respectively
GGG GTA CC
SmaI KpnI-Asp718
GAT CCC GTC GTT TTA CAA CGT CGTGACTGG GAAAACCCTG GCGTTACCCA
Asp Pro Val Val Leu .....LacZ in the +1 frame in pF7
in the 0 frame in pF9
in the -1 frame in pF8
ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC
GAATGGCGCT TTGCCTGGTT TCCGGCACCA GAAGCGGTGC CGGAAA

Plasmid 5D18/3D30XS:

```
1.6 KB of PGK sequences from a HindIII site to ||
     GATCGTACTG TTACTCTCTC TCTTTCAAAC AGAATTGTCC GAATCGTGTG

     ACAACAACAG CCTGTTCTCA CACACTCTTT TCTTCTAACC AAGGGGGTGG

     TTTAGTTTAG TAGAACCTCG TGAAACTTAC ATTTACATAT ATATAAACTT

     GCATAAAATTG GTCAATGCAA GAAATAGATA TTTGGTCTTT TCTAATTCGT

     AGTTTTTCAA GTTCTTAGAT GCTTTCTTTT TCTCTTTTTT ACAGATCATC

     AAGGAAGTAA TTATCTACTT TTTACAACAA AT  || CTA
          ^Start of transcript in this direction -->

               GAATTC ATACAAA ATG ACT TCT AGG
                EcoR1          |^^
                               |Translation Start

                                          5'Pseudoknot
L-A base #     Slippery Site                region
   1956         CAGGGTTTAGGAGTGGTAGGTCTTACGATGCCAGCTGTAATGCCT   2000
                  ^^^^^^     --------> ----->  ^^^^^^<----- <---
                Q  G  L  G  V  V  G  L  T  M  P  A  V  M  P
                   G  F  R  S  G  R  S  Y  D  A  S  C  N  A  Y
                      3'Pseudoknot
                       region
   2001  ACCGGAGAACCTACAGCTGGCGCTGCC  2030
         ----           ^^^^^^
         T  G  E  P  T  A  G  A  A
            R  R  T  Y  S  W  R  C  P

  TAG CTCGAGCCCGGG GTA CC
       XhoI SmaI KpnI=Asp718
          GAT CCC GTC GTT TTA CAA CGT CGTGACTGG GAAAACCCTG GCGTTACCCA
          Asp Pro Val Val Leu .....LacZ in the -1 frame in 5D18/3D30XS

          ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG

          AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC

          GAATGGCGCT TTGCCTGGTT TCCGGCACCA GAAGCGGTGC CGGAAA
```